PCI-SIG ENGINEERING CHANGE NOTICE

TITLE:	MSI-X ECN for PCI Express
DATE:	October 31, 2003
AFFECTED DOCUMENTS:	PCI Express Base Specification 1.0a PCI Express Integrated Devices - Event Collector ECN
SPONSOR:	Joe Cowan; Hewlett-Packard Company

Part I

1. Summary of the Functional Changes

Enhance PCI Express to support MSI-X, as defined in the MSI-X ECN for PCI 2.3. Require PCI Express devices that support interrupts to implement MSI or MSI-X or both.

2. Benefits as a Result of the Changes

PCI Express systems can take advantage of the benefits offered by MSI-X over MSI, notably:

1. Supports a larger maximum number of vectors (2048 instead of 32).
2. Supports independent message address and data for each vector.
3. Supports per-vector masking. (The MSI-X ECN for PCI 2.3 also defines per-vector masking for MSI.)
4. Enables more flexibility when software allocates fewer vectors than hardware requests.

3. Assessment of the Impact

Subsequent pages detail the changes required for the impacted specifications.

4. Analysis of the Hardware Implications

PCI Express devices that generate interrupts are now permitted to implement MSI-X in addition to MSI, or in lieu of MSI.

5. Analysis of the Software Implications

1. If a device implements MSI-X in addition to MSI, any existing software that supports only MSI will not recognize the MSI-X capability, and will operate using MSI. New software that supports MSI-X can take advantage of MSI-X's benefits over MSI.
2. If a device implements MSI-X in lieu of MSI, any existing software that supports only MSI will not recognize the MSI-X capability, and will be unable to configure the device to use messages to signal interrupts. It is believed that little if any such software exists.

Part II

Detailed Description of the changes: PCI Express Base Specification 1.0a

Terms and Acronyms, p. 22, change text as shown:

Message Signaled Interrupt~~,~~ (MSI/MSI-X)	~~An optional feature that enables a device to request service by writing a system-specified DW of data to a system-specified address using a Memory Write semantic Request.~~Two similar but separate mechanisms that enable a device to request service by writing a system-specified DWORD of data to a system-specified address using a Memory Write Request. MSI is defined in PCI 2.3. MSI-X is a separate extension mechanism defined in the MSI-X ECN for PCI 2.3. Compared to MSI, MSI-X supports a larger maximum number of vectors and independent message address and data for each vector.

1.3.2.1. Legacy Endpoint Rules, p. 32, change text as shown:

- A legacy Endpoint is required to support MSI or MSI-X or both, if an interrupt resource is requested. ~~, but~~ If MSI is implemented, a legacy Endpoint is permitted to support either the 32-bit or 64-bit Message Address version of the MSI capability structure.

1.3.2.2. PCI Express Endpoint Rules, p. 33, change text as shown:

- A PCI Express Endpoint is required to support MSI or MSI-X or both, if an interrupt resource is requested. ~~, and~~ If MSI is implemented, a PCI Express Endpoint must support the 64-bit Message Address version of the MSI capability structure.

2.2.7. Memory, I/O, and Configuration Request Rules, p. 61, change text as shown:

Message Signaled Interrupt~~s~~ (MSI/MSI-X~~s~~) mechanisms use Memory Write Requests to represent interrupt Messages (see Section 6.1.4). The Request format used for MSI/MSI-X transactions is identical to the Memory Write Request format defined above, and MSI/MSI-X Requests are indistinguishable from memory writes with regard to ordering, Flow Control, and data integrity.

2.2.8.1. INTx Interrupt Signaling - Rules, p. 63, change text as shown:

~~The~~ A Message Signaled Interrupt (MSI or MSI-X) ~~mechanism~~ is the preferred interrupt signaling mechanism in PCI Express (see Section 6.1). However, in some systems, there may be devices which cannot support the MSI or MSI-X mechanisms. The INTx virtual wire interrupt signaling mechanism is used to support legacy Endpoints and PCI Express/PCI(-X) Bridges in cases where the MSI or MSI-X mechanisms cannot be used. Switches must support this mechanism. The following rules apply to the INTx interrupt signaling mechanism:

6.1. Interrupt and PME Support, p. 259, change text as shown:

The PCI Express interrupt model supports two mechanisms:

- INTx emulation
- Message Signaled Interrupt (MSI/MSI-X) Support.

For legacy compatibility, PCI Express provides a PCI INTx emulation mechanism to signal interrupts to the system interrupt controller (typically part of the Root Complex). This mechanism is compatible with existing PCI software, and provides the same level and type of service as corresponding PCI interrupt signaling mechanism and is independent of system interrupt controller specifics. This legacy compatibility mechanism allows boot device support without requiring complex BIOS-level interrupt configuration/control service stacks. It virtualizes PCI physical interrupt signals by using an in-band signaling mechanism.

In addition to PCI INTx compatible interrupt emulation, PCI Express requires support of MSI or MSI-X or both ~~the Message Signaled Interrupt (MSI) mechanism~~. The PCI Express MSI mechanism is compatible with the MSI capability defined in the PCI 2.3 Specification. The PCI Express MSI-X mechanism is compatible with the MSI-X capability defined in the MSI-X ECN for the PCI 2.3 Specification.

6.1.4. Message Signaled Interrupt (MSI) Support, p. 260, change text as shown:

6.1.4 Message Signaled Interrupt (MSI/MSI-X) Support

~~The Message Signaled Interrupt (MSI) capability is defined in the PCI 2.3 Specification.~~

MSI/MSI-X interrupt support, which is optional for PCI 2.3 devices, is required for PCI Express devices. All PCI Express devices that are capable of generating interrupts must support MSI or MSI-X or both. ~~MSI capable devices~~ The MSI and MSI-X mechanisms deliver interrupts by performing memory write transactions. MSI and MSI-X ~~is~~ are ~~an~~ edge-triggered interrupt mechanisms; neither the PCI 2.3 Specification nor this specification support level-triggered MSI/MSI-X interrupts. Certain PCI devices and their drivers rely on INTx-type level-triggered interrupt behavior (addressed by the PCI Express legacy INTx emulation mechanism). To take advantage of the MSI or MSI-X capability and edge-triggered interrupt semantics, these devices and their drivers may have to be redesigned.

A legacy Endpoint that implements MSI is required to support either the 32-bit or 64-bit Message Address version of the MSI capability structure. A PCI Express Endpoint that implements MSI is required to support the 64-bit Message Address version of the MSI capability structure.

The requestor of an MSI/MSI-X transaction must set the No Snoop and Relaxed Ordering attributes of the Transaction Descriptor to 0.

Note that, unlike INTx emulation Messages, MSI/MSI-X transactions are not restricted to TC0 traffic class.

IMPLEMENTATION NOTE

Synchronization of Data Traffic and Message Signaled Interrupts

MSI/MSI-X transactions are permitted to use the TC that is most appropriate for the device's programming model. This is generally the same TC as is used to transfer data; for legacy I/O, TC0 should be used.

If a device uses more than one TC, it must explicitly ensure that proper synchronization is maintained between data traffic and interrupt Message(s) not using the same TC. Methods for ensuring this synchronization are implementation specific. One option is for a device to issue a zero length Read (as described in Section 2.2.5) using each additional TC used for data traffic prior to issuing the MSI/MSI-X transaction. Other methods are also possible. Note, however, that platform software (e.g., a device driver) is generally only capable of issuing transactions using TC0.

7.5.1.1. Command Register, p. 321, change text as shown:

Table 7-3: Command Register

Bit Location	Register Description	Attributes
2	**Bus Master Enable** – Controls the ability of a PCI Express Endpoint to issue Memory and I/O Read/Write Requests, and the ability of a Root or Switch Port to forward Memory and I/O Read/Write Requests in the upstream direction *Endpoints:* Disabling this bit prevents a PCI Express agent from issuing any Memory or I/O Requests. Note that as MSI/MSI-X interrupt Messages are in-band memory writes, disabling the bus master enable bit disables MSI/MSI-X interrupt Messages as well. Requests other than Memory or I/O Requests are not controlled by this bit. Default value of this field is 0b. This bit is hardwired to 0b if a device does not generate Memory or I/O Requests.	RW

7.7. MSI Capability Structure, p. 333, change text as shown:

7.7 MSI and MSI-X Capability Structures

~~The MSI capability structure is required for all~~ All PCI Express devices that are capable of generating interrupts must implement MSI or MSI-X or both. ~~This~~ The MSI capability is defined in the *PCI Local Bus Specification, Rev. 2.3*. The MSI-X capability and new optional MSI features are defined in the MSI-X ECN for *PCI Local Bus Specification, Rev. 2.3.*

7.8.2. PCI Express Capabilities Register, p. 336, change text as shown:

Table 7-10: PCI Express Capabilities Register

Bit Location	Register Description	Attributes
…	…	…
13:9	**Interrupt Message Number** – ~~If this function is allocated more than one MSI interrupt number, this register is required to contain the offset between the base Message Data and the MSI Message that is~~ This register must indicate which MSI/MSI-X vector is used for the interrupt message generated when any of the status bits in either the Slot Status register or the Root Port Status register of this capability structure are set. For MSI, the value in this register indicates the offset between the base Message Data and the interrupt message that is generated. Hardware is required to update this field so that it is correct if the number of MSI Messages assigned to the device changes when software writes to the Multiple Message Enable field in the MSI Message Control register. For MSI-X, the value in this register indicates which MSI-X Table entry is used to generate the interrupt message. The entry must be one of the first 32 entries even if the function implements more than 32 entries. For a given MSI-X implementation, the entry must remain constant. If both MSI and MSI-X are implemented, they are permitted to use different vectors, though software is permitted to enable only one mechanism at a time. If MSI-X is enabled, the value in this register must indicate the vector for MSI-X. If MSI is enabled or neither is enabled, the value in this register must indicate the vector for MSI. If software enables both MSI and MSI-X at the same time, the value in this register is undefined.	RO

7.10.10. Root Error Status Register, p. 376, change text as shown:

Table 7-32: Root Error Status Register

Bit Location	Register Description	Attributes
…	…	…
31:27	**Advanced Error Interrupt Message Number** —~~If this function is allocated more than one MSI interrupt number, this register is required to contain the offset between the base Message Data and the MSI Message that is~~ This register must indicate which MSI/MSI-X vector is used for the interrupt message generated when any of the status bits of this capability are set. For MSI, the value in this register indicates the offset between the base Message Data and the interrupt message that is generated. Hardware is required to update this field so that it is correct if the number of MSI Messages assigned to the device changes when software writes to the Multiple Message Enable field in the MSI Message Control register. For MSI-X, the value in this register indicates which MSI-X Table entry is used to generate the interrupt message. The entry must be one of the first 32 entries even if the function implements more than 32 entries. For a given MSI-X implementation, the entry must remain constant. If both MSI and MSI-X are implemented, they are permitted to use different vectors, though software is permitted to enable only one mechanism at a time. If MSI-X is enabled, the value in this register must indicate the vector for MSI-X. If MSI is enabled or neither is enabled, the value in this register must indicate the vector for MSI. If software enables both MSI and MSI-X at the same time, the value in this register is undefined.	RO

1.3.2.3. Root Complex Integrated Endpoint Rules, change text as shown:

- A Root Complex Integrated Endpoint is required to support MSI or MSI-X or both, if an interrupt resource is requested. , ~~but~~If MSI is implemented, a Root Complex Integrated Endpoint is permitted to support either the 32-bit or 64-bit Message Address version of the MSI capability structure.